A multigrid tour of ocean DA

well... a short one

A. Vidard and friends

Motivations

There several (possibly conflicting) reasons for using multi grid approaches when dealing with minimization:

- to reduce the computing cost,
- to improve the rate of convergence,
- to deal with non linearities.

However it may not be trivial for the ocean

- different behaviour at different resolution,
- requires handling of complex boundaries,
- can be technically challenging

Outline

01

Multiscale, Nested, Coarse...

- different ”scales” are assimilated separately
- In general 2 assimilation kernels are used
- not all of them use multiple grids, but
    - it can be used to acheive scale separation
    - it can be used to save time assimilating large scales

Anthony's correlation operator $\mathbf{C}^* = \mathbf{N}^{1/2}\mathbf{L}^{1/2}\mathbf{W}_f^{-1}\mathbf{L}^{T/2}\mathbf{N}^{1/2}$ can be

approximated by:

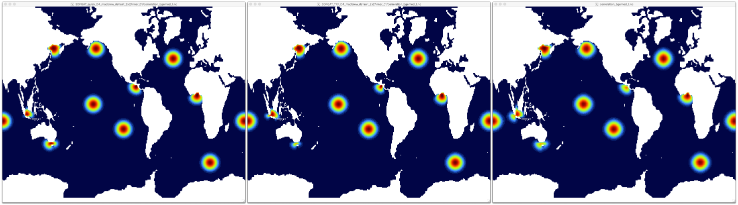
$$\mathbf{C} = \mathbf{N}^{1/2}\mathbf{T}\mathbf{L}_c^{1/2}\mathbf{W}_c^{-1}\mathbf{L}_c^{T/2}\mathbf{T}^T\mathbf{N}^{1/2}$$

Remarks:

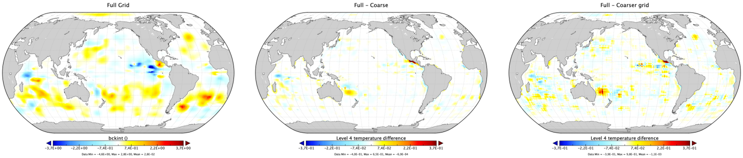
- Other components of **B** (e.g balance) stay at fine resolution
- The transfer from fine to coarse is $\mathbf{T}^T = \mathbf{E}^T\mathbf{F}^T$ where **F** is a smoothing filter .
- The impact of $\mathbf{F}^T$ is probably minor since it is followed by the application of **L** which will do large-scale smoothing.
- The impact of **F** should be more important since **E** will produce step-like fields on the fine grid, that will need to be smoothed accordingly.

If the prescribed decorrelation scales are large enough, the approximation is reasonable. In examples below it is set to 4 times the fine resolution grid cell size.

effect of **C** on isolated impulses is well represented with a coarsening ratio of 2, and tend to be diamond shaped with 3

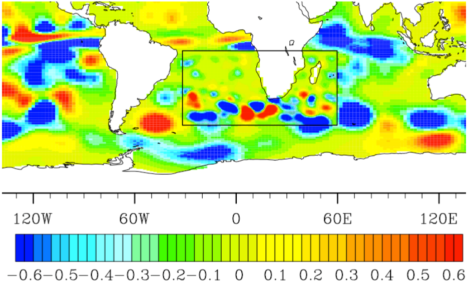


By eye it is difficult to spot the difference on the increments, but the difference shows isolated location where the error can reach 10%



Local High resolution nested grids are used in the forecast model (Simon et al, 2010)

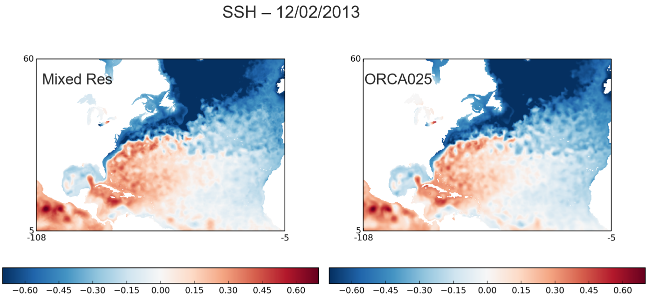
- No significant modification to the assimilation scheme
- Chalenges similar to that of coupled data assimilation



Quite often, in order to save computing time, the analysis (and/or the ensemble) runs at coarser resolution.

Few things one has to ask himself:

- Is the coarse dynamics representative enough?
- How to 'synchronize' fine and coarse models?
- What about high resolution observations?
- Why not doing everything at coarse resolution?



02

Multigrid proper

lets consider the semi discretised model

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} & = & \mathcal{M}(\mathbf{x}), \quad t \in [0, T], \\ \mathbf{x}(0) & = & \mathbf{x}_0, \end{cases} \tag{1}$$

The classical 4D-Var algorithm leads to solve

$$\min_{\mathbf{x}_0 \in \mathbb{R}^n} J(\mathbf{x}_0) = \frac{1}{2} \int_0^T \| \mathbf{y}(t) - \mathcal{H}[\mathbf{x}(\mathbf{x}_0, t)] \|_{\mathcal{O}}^2 \, dt \ + \frac{1}{2} \| \mathbf{x}_0 - \mathbf{x}^b \|_{\mathcal{X}}^2,$$

(with
$\| \mathbf{x}_0 - \mathbf{x}^b \|_{\mathcal{X}}^2 := (\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b))$
and
$\| \mathbf{y}(t) - \mathcal{H}[\mathbf{x}(\mathbf{x}_0, t)] \|_{\mathcal{O}}^2 = (\mathbf{y}(t) - \mathcal{H}[\mathbf{x}(\mathbf{x}_0, t)])^T \mathbf{R}^{-1} (\mathbf{y}(t) - \mathcal{H}[\mathbf{x}(\mathbf{x}_0, t)])$

The 'outer' cost function:

$$J(\delta \mathbf{x}_0) = \frac{1}{2}\int_0^T \| \mathbf{y}(t) - \mathcal{H}[\mathcal{M}(\mathbf{x}^b + \delta \mathbf{x}_0, t)] \|_{\mathcal{O}}^2 \, dt \; + \frac{1}{2} \| \delta \mathbf{x}_0 \|_{\mathcal{X}}^2,$$

That is then minimised using a solution algorithm based on incremental 4D-Var (Gauss-Newton): (courtier *et al.* 1994)

$$\min_{\delta \mathbf{x}_0^{(k)} \in \mathbb{R}^n} J^{(k)}(\delta \mathbf{x}_0^{(k)}) = \frac{1}{2}\int_0^T \| \mathbf{H}^{(k)}\mathbf{M}^{(k)}\delta \mathbf{x}_0^{(k)} - \delta \mathbf{y}^{(k-1)}(t) \|_{\mathcal{O}}^2 \, dt$$
$$+ \frac{1}{2} \| \delta \mathbf{x}_0^{(k)} - \delta \mathbf{x}^{b,(k-1)} \|_{\mathcal{X}}^2,$$

- $\delta \mathbf{x}^{b,(k-1)} = \mathbf{x}^b - \mathbf{x}_0^{k-1}$ and $\delta \mathbf{y}^{(k-1)} = \delta \mathbf{y}^{(k-1)} - \mathcal{H}[\mathcal{M}(\mathbf{x}_0^{(k-1)}, t)]$
- **Inner problem**: $\delta \mathbf{x}_0^{(k)}$ found using CG minimizer
- **Outer problem**: very few iterations k are affordable in practice.
- Setting $\mathbf{M}^{k-1} = \mathbf{I}$ gives the (much cheaper!) 3D-FGAT algorithm (widely used in ocean DA).

Under the right hypotheses, GN converges locally to $\delta \mathbf{x}_0^*$ the minimum of J.

This can be extended to the multi-incremental 4D-Var (Veersé and Thépaut, 1998) aka Perturbed GN where the inner problem is solved at a lower resolution $\delta\mathbf{x}_{0,c}^{(k)} = I_f^c \delta\mathbf{x}_0^{(k)}$ with possibly simplified physics ; Same Non linear problem:

$$J(\delta\mathbf{x}_{0,f}) = \frac{1}{2}\int_0^T \parallel \mathbf{y}(t) - \mathcal{H}_f[\mathcal{M}_f(\mathbf{x}_f^b + \delta\mathbf{x}_{0,f}, t)] \parallel_{\mathcal{O}}^2 \, dt \; + \frac{1}{2} \parallel \delta\mathbf{x}_{0,f} \parallel_{\mathcal{X}}^2,$$

the inner loop becomes

$$\min_{\delta\mathbf{x}_{0,c}^k \in \mathbb{R}^m} J_c^{(k)}(\delta\mathbf{x}_{0,c}^{(k)}) = \frac{1}{2}\int_0^T \parallel \mathbf{H}_c^{(k)}\mathbf{M}_c^{(k)}\delta\mathbf{x}_{0,c}^{(k)} - \delta\mathbf{y}^{(k-1)}(t) \parallel_{\mathcal{O}}^2 \, dt$$
$$+ \frac{1}{2} \parallel \delta\mathbf{x}_{0,c}^{(k)} - I_f^c \delta\mathbf{x}_f^{b,(k-1)} \parallel_{\mathcal{X}}^2,$$

This can be extended to the multi-incremental 4D-Var (Veersé and Thépaut, 1998) aka Perturbed GN where the inner problem is solved at a lower resolution $\delta\mathbf{x}_{0,c}^{(k)} = I_f^c \delta\mathbf{x}_0^{(k)}$ with possibly simplified physics ; Same Non linear problem:

$$J(\delta\mathbf{x}_{0,f}) = \frac{1}{2}\int_0^T \parallel \mathbf{y}(t) - \mathcal{H}_f[\mathcal{M}_f(\mathbf{x}_f^b + \delta\mathbf{x}_{0,f}, t)] \parallel_{\mathcal{O}}^2 dt \; + \frac{1}{2} \parallel \delta\mathbf{x}_{0,f} \parallel_{\mathcal{X}}^2,$$

the inner loop becomes

$$\min_{\delta\mathbf{x}_{0,c}^k \in \mathbb{R}^m} J_c^{(k)}(\delta\mathbf{x}_{0,c}^{(k)}) \qquad = \qquad J_c^{o(k)}(\delta\mathbf{x}_{0,c}^{(k)}) \quad + \quad J_c^{b(k)}(\delta\mathbf{x}_{0,c}^{(k)}),$$

This can be extended to the multi-incremental 4D-Var (Veersé and Thépaut, 1998) aka Perturbed GN where the inner problem is solved at a lower resolution $\delta\mathbf{x}_{0,c}^{(k)} = I_f^c \delta\mathbf{x}_0^{(k)}$ with possibly simplified physics ; Same Non linear problem:

$$J(\delta\mathbf{x}_{0,f}) = \frac{1}{2}\int_0^T \parallel \mathbf{y}(t) - \mathcal{H}_f[\mathcal{M}_f(\mathbf{x}_f^b + \delta\mathbf{x}_{0,f}, t)] \parallel_{\mathcal{O}}^2 \, dt \; + \frac{1}{2} \parallel \delta\mathbf{x}_{0,f} \parallel_{\mathcal{X}}^2,$$

the inner loop becomes

$$\min_{\delta\mathbf{x}_{0,c}^k \in \mathbb{R}^m} J_c^{(k)}(\delta\mathbf{x}_{0,c}^{(k)}) \quad = \quad J_c^{o(k)}(\delta\mathbf{x}_{0,c}^{(k)}) \quad + \quad J_c^{b(k)}(\delta\mathbf{x}_{0,c}^{(k)}),$$

This algorithm converges locally toward $\delta\tilde{\mathbf{x}}_0^* \neq \delta\mathbf{x}_0^*$ (Gratton *et al.* 2007) , approximation that can degrade the analysis (Trémolet 2007) depending on how $\mathbf{H}_c\mathbf{M}_c$ approximates $\mathcal{H}_f(\mathcal{M}_f(.))$.

Multigrid schemes 2: GN Multi Grid

(Neveu 2011) Proposed to use a FAS scheme to solve the inner loops instead (GN-MG) where the outer iterations remain:

$$J(\delta\mathbf{x}_{0,f}) = \frac{1}{2}\int_0^T \parallel \mathbf{y}(t) - \mathcal{H}_f[\mathcal{M}_f(\mathbf{x}_f^b + \delta\mathbf{x}_{0,f}, t)] \parallel_{\mathcal{O}}^2 \, dt \; + \frac{1}{2} \parallel \delta\mathbf{x}_{0,f} \parallel_{\mathcal{X}}^2,$$

the inner loop becomes:

1. look for (1-2 iterations)

$$\min_{\delta\mathbf{x}_{0,f}^k \in \mathbb{R}^n} J_f^{(k)}(\delta\mathbf{x}_{0,f}^{(k)}) = J_f^{o(k)}(\delta\mathbf{x}_{0,f}^k) + J_f^{b(k)}(\delta\mathbf{x}_{0,f}^k),$$

(Neveu 2011) Proposed to use a FAS scheme to solve the inner loops instead (GN-MG) where the outer iterations remain:

$$J(\delta \mathbf{x}_{0,f}) = \frac{1}{2}\int_0^T \parallel \mathbf{y}(t) - \mathcal{H}_f[\mathcal{M}_f(\mathbf{x}_f^b + \delta \mathbf{x}_{0,f}, t)] \parallel_{\mathcal{O}}^2 \, dt \; + \frac{1}{2} \parallel \delta \mathbf{x}_{0,f} \parallel_{\mathcal{X}}^2,$$

the inner loop becomes:

1. look for (1-2 iterations)

$$\min_{\delta \mathbf{x}_{0,f}^k \in \mathbb{R}^n} J_f^{(k)}(\delta \mathbf{x}_{0,f}^{(k)}) = J_f^{o(k)}(\delta \mathbf{x}_{0,f}^k) + J_f^{b(k)}(\delta \mathbf{x}_{0,f}^k),$$

2. from $\delta \mathbf{x}_{0,f}^{*(k)}$

$$\min_{\delta \mathbf{x}_{0,c}^k \in \mathbb{R}^m} \tilde{J}_c^{(k)}(\delta \mathbf{x}_{0,c}^{(k)}) = J_c^{(k)}(\delta \mathbf{x}_{0,c}^k) - \underbrace{\left\langle \nabla J^c(I_f^c \delta \mathbf{x}_{0,f}^{*(k)}) - I_f^c \nabla J^f(\delta \mathbf{x}_{0,f}^{*(k)}), \delta \mathbf{x}_{0,c}^k \right\rangle}_{J_{FAS}},$$

(Neveu 2011) Proposed to use a FAS scheme to solve the inner loops instead (GN-MG) where the outer iterations remain:

$$J(\delta \mathbf{x}_{0,f}) = \frac{1}{2}\int_0^T \parallel \mathbf{y}(t) - \mathcal{H}_f[\mathcal{M}_f(\mathbf{x}_f^b + \delta \mathbf{x}_{0,f}, t)] \parallel_{\mathcal{O}}^2 \, dt \; + \frac{1}{2} \parallel \delta \mathbf{x}_{0,f} \parallel_{\mathcal{X}}^2,$$

the inner loop becomes:

1. look for (1-2 iterations)

$$\min_{\delta \mathbf{x}_{0,f}^k \in \mathbb{R}^n} J_f^{(k)}(\delta \mathbf{x}_{0,f}^{(k)}) = J_f^{o(k)}(\delta \mathbf{x}_{0,f}^k) + J_f^{b(k)}(\delta \mathbf{x}_{0,f}^k),$$

2. from $\delta \mathbf{x}_{0,f}^{*(k)}$

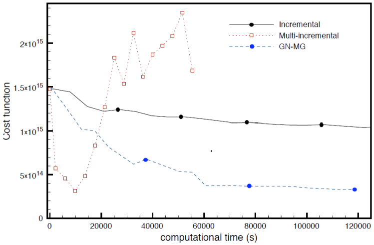
$$\min_{\delta \mathbf{x}_{0,c}^k \in \mathbb{R}^m} \tilde{J}_c^{(k)}(\delta \mathbf{x}_{0,c}^{(k)}) = J_c^{(k)}(\delta \mathbf{x}_{0,c}^k) - \underbrace{\left\langle \nabla J^c(I_f^c \delta \mathbf{x}_{0,f}^{*(k)}) - I_f^c \nabla J^f(\delta \mathbf{x}_{0,f}^{*(k)}), \delta \mathbf{x}_{0,c}^k \right\rangle}_{J_{FAS}},$$

3. update $\delta \mathbf{x}_{0,f}^k = \delta \mathbf{x}_{0,f}^{*k} + I_c^f(\delta \mathbf{x}_{0,c}^{*k} - I_f^c \delta \mathbf{x}_{0,f}^{*k})$ and search again (1-2 iterations) for

$$\min_{\delta \mathbf{x}_{0,f}^k \in \mathbb{R}^n} J_f^{(k)}(\delta \mathbf{x}_{0,f}^{(k)}).$$

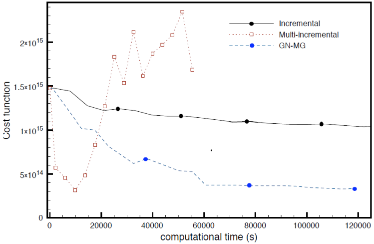
At last we have a multigrid methods that solve the original problem while being (hopefully) efficient. Let's see how it behave in practice:

- 2D Shallow Water model
- 2 grids, refinement ratio 3
- 4Dvar
- diffusion+balance B

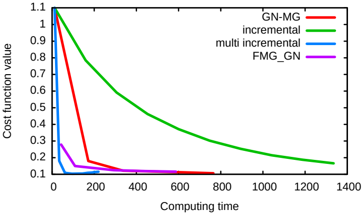


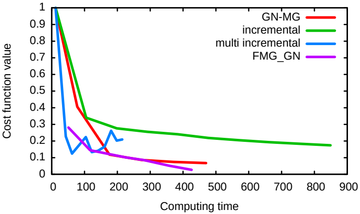
At last we have a multigrid methods that solve the original problem while being (hopefully) efficient. Let's see how it behave in practice:

- 2D Shallow Water model
- 2 grids, refinement ratio 3
- 4Dvar
- diffusion+balance B



Same kind of stuff with Nemovar:





Linear 1D model

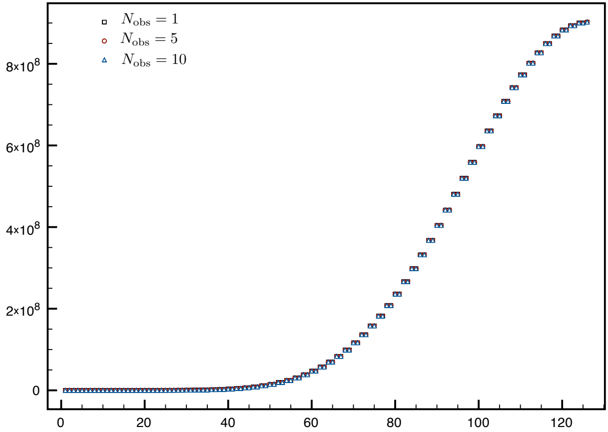
$$\left\{ \frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} \quad = \quad 0, u(0) = u_0 \right. \tag{2}$$

and classical var function

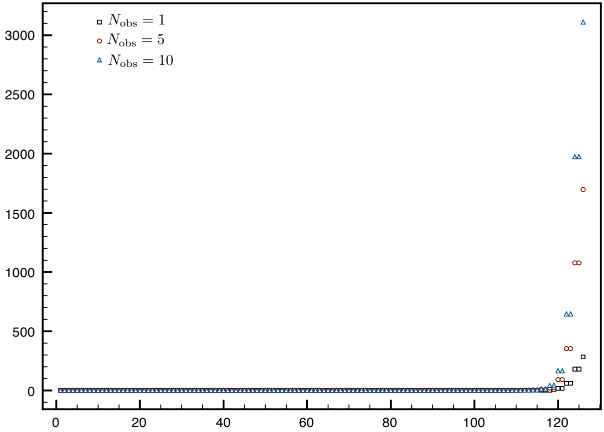
$$J(u_0) = \frac{1}{2}\|u_0 - u^b\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{i=1}^{N_{obs}} \|H_i u(t_i) - y_i\|^2_{\mathbf{R}_i^{-1}}$$

With **B** the implicit diffusion based operator and **R** a diagonal matrix.
let **A** be the Hessian of J and study its eigenvalues

Of the non-preconditionned Hessian $\mathbf{A} = \mathbf{B}^{-1} + \mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{M}$



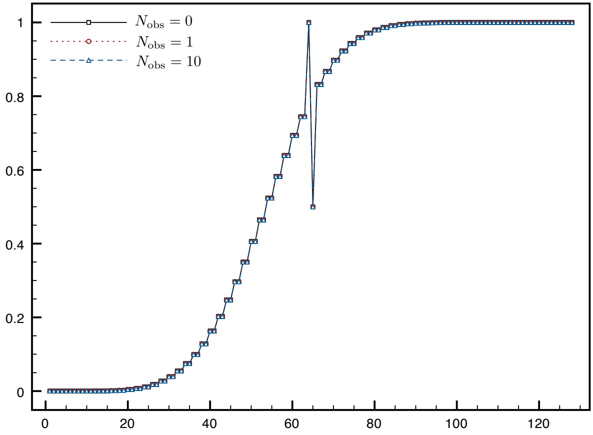
Of the preconditionned Hessian $\mathbf{A} = \mathbf{I} + \mathbf{B}^{T/2}\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{M}\mathbf{B}^{1/2}$



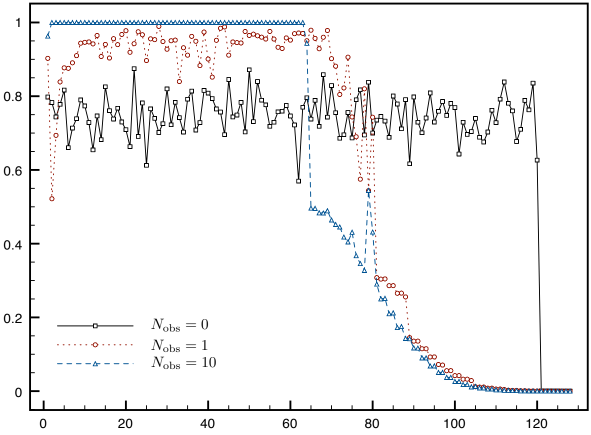
For multi grid methods to be efficient one sought property is ellipticity, i.e. small eigenvalues correspond to large scales. Let define the scale operator:

$$S(v) = \frac{\|(I - I_c^f I_f^c)v\|_2}{\|v\|_2}$$

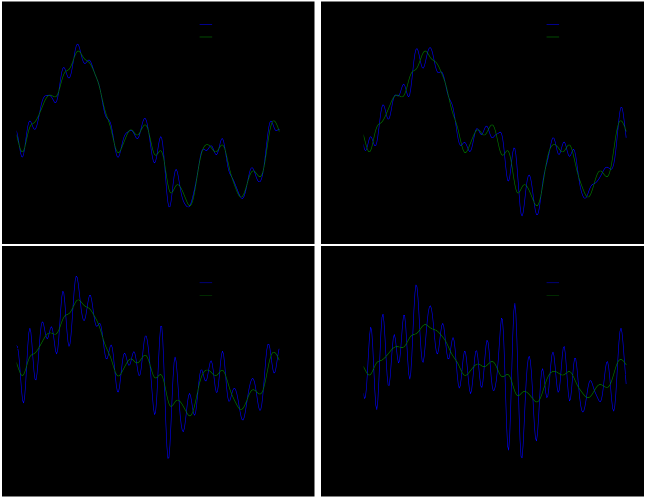
that is close to one at small scales and small at large scales.


$N_{\mathrm{obs}} = 0$
$N_{\mathrm{obs}} = 1$
$N_{\mathrm{obs}} = 10$
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120

Ellipticity



the non ellipticity of the problem makes that the small scales of the errors are not quickly reduced by the minimization method and this may lead, through aliasing on the coarse grid, to a divergence of the multigrid cycles.



- Give up **B** preconditioning (and get rid of Anthony)
  - > we'll still need the inverse
- Do more sweeps at fine resolution (that is what I did)
  - > It is getting too expensive
- Use the multi grid process within the preconditioning

$$(\mathrm{P}): \qquad I_c^f \mathbf{A}_c^{-1} I_f^c + (\mathbf{I}_f - P_c^f I_f^c)^T (\mathbf{I} - P_c^f I_f^c)$$

or

$$(\mathrm{NP}): \qquad I_c^f \mathbf{A}_c^{-1} I_f^c + (\mathbf{I}_f - P_c^f I_f^c)^T \mathbf{B}_f (\mathbf{I} - P_c^f I_f^c)$$